# Interactive Linearisation in Hypertext Information Access

**Thomas Tan**
School of Computing Science, Middlesex University
Bounds Green Road, London N11 2NQ, UK
+ 44 (0)181 362 6714
thomas12@mdx.ac.uk

**ABSTRACT**
SearchLineariser makes use of information contained in hypertext link structure for the linearisation, display and interactive usage of web-site search results. A highly manipulable outline is used to display the linearised and 'editable' document within the web environment. The design and implementation of the system is outlined and possible extensions are discussed.

**Keywords**
Information Access, lost in hyperspace, search linearisation, editable search results, hypertext linearisation, algorithms, links, outline, trees, graphs.

**INTRODUCTION**
Many web documents are comprised of separate web pages that are hyperlinked together. This fact is ignored by many existing web searching methods, which treat each individual web page as a complete document. While these methods are suitable when dealing with traditional "well controlled collections" [1], they do not take into account the many varied forms that real world web pages can come in [2]. Very frequently, the result of a search will contain hits that point the user to what is, arguably,

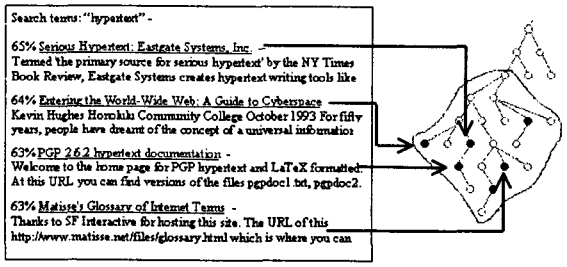


**Figure 1**

only a part of what should be considered as a complete document. The tree structure in Figure 1 represents a part of a searched site, with the enclosed region depicting a web document.

Furthermore, when faced with existing results displays, such as that which is shown in Figure 1, the user is forced to replace and re-display the search results repeatedly when accessing hit documents and the documents linked from those documents. This exposes the user to the risk of disorientation and lostness [3] associated with hypertext navigation. Existing web searching methods and results displays do not consider the effects of the hypertext environment in their design.



**SEARCH LINEARISATION**
This research combines hypertext linearisation with information access. Typically, hypertext linearisation requires the specification of a source page from which to begin the linearisation process [5, 6, 7]. The search linearisation process begins from a user query. This is in contrast to hypertext linearisation, which assumes that the user already knows where the desired information is. Technically, search linearisation poses a quite different challenge. A search returns one or more hits which can be treated as source, sink or intermediate pages. On the simplest level, a straight forward linearisation may not even be possible since hits can exist in different parts or sub-trees of the site. Furthermore, since the current context is searching, the question of whether intermediate pages that are not hits should be included in the final result set arises.

Therefore, the requirement is to formulate a search-linearisation algorithm that derives useful meta-information *inherent* in hypertext by combining graph and tree processing techniques, IR based statistical content-analysis, and other properties of web-hypertext to determine the best linearisation order. The goal is to maintain an optimum level of coherency within the linearised document, given that the linearised document may include pages that are not contained in the original hit set.

Displaying web pages as a single linearised document has many advantages. Firstly, a single linearised document within a scrollable window is more navigable and familiar to users than hypertext traversal. Secondly, the overall context of a document is preserved when information is accessed linearly. The length of the document is indicated by features such as, for instance, the position of the scrollbox in relation to the scrollbar. More importantly, there is a sense of progress when accessing information in a linear fashion, since it is possible to know how much one has read and how much remains. The navigational problems associated with hypertext do not arise when information is accessed linearly [5]. Finally, a linearised document is also more portable since it can be saved or printed and viewed offline.

One problem with linear documents is that they do not scale well onto a standard display screen. Additionally, studies show that users need to understand the nature of

the results relative to the rest of the site [4]. To address these issues, an outline mechanism is used to provide a table-of-contents view of the resulting document. The outline is designed to be highly manipulable, allowing the user to rearrange or delete unwanted hits from the linearised document. Effectively, this turns a search results page into a word processor-like environment.

### SearchLineariser

SearchLineariser is a client/server application that runs in the web environment. To begin, SearchLineariser creates a search index and a tree index by traversing within a given URL's web site. The search index functions as a traditional Information Retrieval index which is used to generate the initial search hits. The tree index is used to remember the web site's hypertext structure. Initially, the tree index is stored in a data structure in the form of a graph. The graph structure is converted into a tree by breaking all the links to previously encountered pages during the traversal. The conversion process is necessary to create an acyclic structure which has a finite number of start and end nodes.

When given a query during operational searching, SearchLineariser first processes the search index to obtain a list of hits. This operation is similar to that performed by most existing search systems. However, unlike existing systems, which display the list of hits to the user and terminates, SearchLineariser displays the hits in an interactive outline. In the outline, the link text of each hit is indented to reflect its topographical location within the original web site. This is accomplished by matching each hit to its corresponding entry in the tree index. The display also reflects each hit's relationship to other hits. Figure 2 shows the resulting outline of a search (browser not shown).

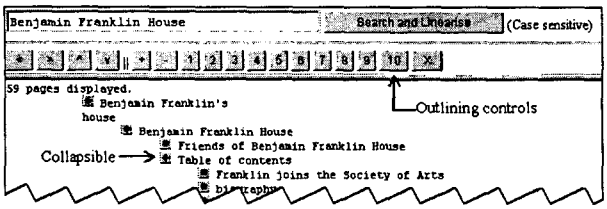


**Figure 2**

The Outline is interactive. "Collapse" and "Expand" manipulation icons are prepended to each hit's link-text. The icons allow the user to collapse or expand sections of the linearised document to hide or show pages. This way, the risk of getting "lost in hyperspace" is minimised, if not removed, since the user is able to view the hits without traversing any links, so to speak.

The Outlining controls are modelled closely to the outlining features found in modern word processing applications. They allow the user to control the levels of viewable pages. The buttons labelled "1" to "10" selects the number of levels to display. The "+" and "-" buttons increase and decrease, respectively, the level to display in single steps. The resulting outline of a fresh search displays all levels. The same happens when the "ALL" button is depressed. The buttons: "<" (shift left), ">" (shift right), "^" (shift up), and "v" (shift down), is used to move highlighted headings in the respective directions.

## CONCLUSIONS AND FURTHER WORK

A recent study showed that 85% of users use search engines to locate information, and that several search engines consistently rank among the top ten sites accessed on the web [8]. Here, I have argued that existing search engine designs effect portals to lostness, and because of their popularity, are therefore a significant source of usability problems to information accessibility and general web usage. Further, I have shown that the nature of search results presentations may not have to remain purely presentational. The ideas and motivations behind the linearisation and display of 'editable' web-site search results have been introduced. The design and implementation of the SearchLineariser[1] has been presented. Ongoing work is extending the implementation to include the ability of displaying web-page level details as well as the already available hypertext tree-level details. Other areas for further work include performing a user study to compare this form of results presentation to existing results presentations such as the ubiquitous list-type results used by commercial search engines.

## ACKNOWLEDGEMENTS

I am grateful to The Royal Society for the encouragement of Arts, Manufactures & Commerce[2] for kindly providing a copy of their web-site for this research.

This work is supported by a postgraduate studentship from the School of Computing Science, Middlesex University.

## REFERENCES

1. Brin, S., Page L : 1998, 'The Anatomy of a Large-Scale Hypertextual Web Search Engine', WWW7. Proceedings of WWW7, 1998.
2. Brown, E.W., Smeaton, A.F.: 1998, "Hypertext Information Retrieval for the Web", ACM SIGIR'98 Post-Conference Workshop on "Hypertext Information Retrieval for the Web.
3. Conklin, J.: 1987, 'Hypertext: an introduction and survey. IEEE Computer, Sept, pp17-41.
4. Nielson, J.: 1997, 'Search and you may find', in Jakob Nielsen's Alertbox for July, 15, 1997. http://www.useit.com/alertbox/9707b.html.
5. Bench-Capon, T.J.M., Dunne, R.J., Staniford, G,: 1992, 'Linearising Hypertext through Target Graph Specification', Database and Expert Applications, no.3, pp.173-178, 1992.
6. Gagnon, M., Dagenais, M.: 1997, 'From a Web to a linear HTML document', WWW6. In Proceedings of WWW6, 1997.
7. Sharples, Mike, Goodlet, James,: 1994, 'A comparison of algorithms for hypertext notes network linearisation', Int. J. Human-Computer Studies, vol. 40, 1994, pp. 727-752.
8. Lawrence, S., Giles, C.L.: 1999, 'Accessibility of information on the web', Nature, 8 July 1999.

[1] http://www.thomas.mdx.ac.uk/SearchLinearliserHome

[2] http://www.rsa.org.uk